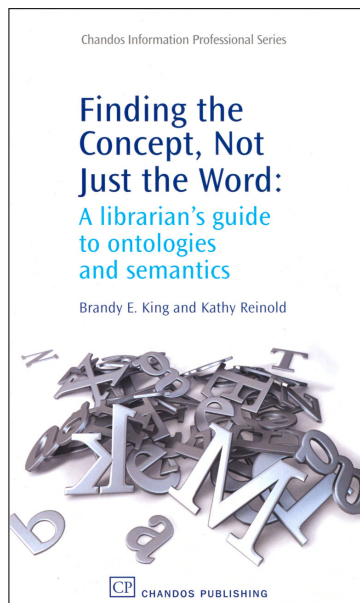


■ **Brandy E. King, Kathy Reinold: Finding the Concept, Not Just the Word. A librarian's guide to ontologies and semantics (= Chandos information professional series), Oxford, UK: Chandos, 2008. xxi, 202 S.**
ISBN 978-1-84334-318-9 (br., £ 39,95), 978-1-84334-319-6 (geb., £ 59,95)

Die Autorinnen dieses Buches kommen aus den USA und sind in den Bereichen Medizinbibliothek (King) und Informatik (Reinold) tätig. Wie dem Vorwort zu entnehmen ist, liegt ihrer gemeinsamen Publikation ein berufliches Kooperationsprojekt zugrunde, als Folge dessen der Wunsch entstand, eine verständliche Einführung in das Gebiet Ontologien und semantische Technologien zu verfassen. Dass man sich dabei laut Untertitel explizit an die bibliothekarische Fachwelt wendet, wird im Buchinneren allerdings nicht wirklich eingelöst, zumal der Text eigentlich von jedermann gelesen werden kann. Vielleicht wollte man aber Bibliothekare und Informationsspezialisten deshalb gezielt ansprechen, da für sie, wie im Text mehrmals betont wird, bei der Umsetzung von Suchanfragen in Suchstrategien und den verwendeten Retrievalinstrumenten bis dato das Paradigma des *text matching* gilt. Letzteres umso mehr (so könnte man ergänzen), da in der Bibliothekswelt gegenwärtig neue und nicht gerade wohlfeile Abfrageinstrumente eingeführt werden, die der durch Google und E-Commerce verblendeten Klientel vorgaukeln, semantische Suchprobleme mehr oder weniger allein durch die Kombination von Stichwortsuche und Ranking lösen zu können.



Dagegen setzen die Autorinnen im einführenden Kapitel die Frage: „... what if the search engine understood the concepts in the question, rather than just matching the text?“ und meinen weiter: „This is all possible with the use of ontologies and semantic technologies.“ (beides p.xix). Ist dies nur blauäugiger Optimismus oder stecken dahinter handfeste, alltagstaugliche Alternativen? Die Antwort, so sei vorweggenommen, ist ein klares „Ja“. King und Reinold versuchen jedenfalls, ihr Anliegen in fünf Hauptabschnitten darzulegen und ihrer Darstellung dabei auch den Charakter eines Lehrtextes zu geben. Letzteres gelingt gar nicht schlecht, zumal das Buch in verständlicher Sprache geschrieben ist, am Ende jedes der sieben Kapitel der beiden ersten Hauptabschnitte ein quasi wiederholender Abschnitt („Key concepts“) angefügt ist und auch sonst immer wieder zusammenfassende oder vergleichende Tabellen und Grafiken den Text ergänzen.

Der erste und längste Teil (64 Seiten) ist mit *Understanding ontologies* überschrieben und enthält vier Kapitel, die sich mit folgenden Themen beschäftigen:

- Was sind Ontologien und wie unterscheiden sie sich von anderen Mitteln der Wissensrepräsentation? Obgleich die Definition von Gruber (1993) bzw. jene des W3C (2004) angeführt werden, legen die Autorinnen für das vorliegende Buch ausdrücklich fest, eine Ontologie sei „a collection of concepts, arranged in a hierarchy of categories, combined with the relationships between those concepts, in order to reflect the vocabulary of an area of knowledge.“ (p. 8). Dies ist nun so allgemein gehalten, dass etwa Thesauren und Klassifikationssysteme locker darunterfallen würden. Doch das ist hier nicht gemeint, denn nach King und Reinold unterscheiden sich letztere durch die Absenz von Definitionen und eine weit weniger reichhaltige Ausstattung mit Relationen verschiedenster Art (worüber es sich allerdings durchaus streiten liesse).
- Wie können im Gegensatz zu herkömmlichen Retrievalinstrumenten Sucherfolg und Relevanz durch Verwendung von Ontologien bestimmt bzw. verbessert werden? Hier werden zunächst semantischen Techniken wie NLP (natural language processing), semantische Analyse, Referencing (Zuordnung von Begriffen/Synonymen zu Ontologien und Ersetzen von Begriffen durch Identifikatoren bzw. Vorzugsbegriffe) und Ranking auf Basis der wichtigsten Begriffe aus den Benutzeranfragen vorgestellt. Sodann erfolgt die Darstellung von Aufbau und Ablauf in einem semantischen Suchsystem (Analyse the natural language; Map the words/phrases to concepts in the

ontology; Create a semantic index; Query; Rank the results) unter Verweis auf drei bereits existierende Lösungen.

- Wie erstellt man eine Ontologie? Dafür geben die Autorinnen anhand eines Beispiels aus dem gastronomischen Bereich eine entsprechend kochbuchartig anmutende Anleitung, die neun Schritte umfasst: 1. Decide how your ontology will be used; 2. Gather terms related to your domain; 3. Create a taxonomy (gemeint ist hier eine Hierarchie von Kategorien); 4. Add more terms to each category; 5. Define terms (im domänenspezifischen Sinn!); 6. Add synonyms; 7. Define relationships and use them to connect terms; 8. Verify your ontology; 9. Revise. Dies klingt natürlich wie eine traditionelle Anleitung zur Konstruktion einer Fachsystematik oder eines Fachthesaurus und unterscheidet sich hievon eigentlich nur durch den Verweis auf die darüber hinausgehenden Arten von Beziehungen (actions; locations; conceptual connections), die im siebenten Schritt zu definieren sind.
- Wie erstellt man eine Ontologie auf der Basis bereits existierender Komponenten (Taxonomie, Thesaurus, vorhandene Ontologie)? Hier wird, unter Rückgriff auf die eben dargestellten neun Schritte und mitunter ziemlich repetitiv, auf die Unterschiede zu „from scratch“ eingegangen. Dass in beiden Kapiteln als Beispieldomäne keine wissenschaftliche, sondern eine solche aus dem Alltag herangezogen wird, mag zwar das Verständnis erleichtern, erschwert es aber, sich die Übertragung auf wissenschaftliche Themenbereiche vorzustellen.

Der zweite Hauptteil (32 Seiten) ist der Einführung in die semantischen Technologien gewidmet. Darunter verstehen King und Reinold „any automated software tool that processes information based on an interpretation of meaning“ (p. 65). Diese Interpretation bedeutet hauptsächlich die Fähigkeit zur Erkennung unterschiedlicher Wege für die Identifizierung derselben Vorstellung bzw. desselben Begriffs. Andere Technologien gehen darüber hinaus und ermöglichen bspw. die Ableitung neuer Fakten auf der Basis existierender Informationen. Die drei Kapitel dieses Abschnittes berichten über:

- Natürliche Spracherkennung (NLP), wobei die beiden grundsätzlichen Ansätze vorgestellt werden, d.h. der grammatik-basierte und der auf Mustererkennung und statistischer Analyse beruhende Ansatz; ein Trend zur Kombination beider wird konstatiert.
- Die Verwendung von Metadaten zur Anreicherung mit semantischer Information. Hier wird kurz auf strukturierte Sprachen wie SGML und XML, sowie auf „metadata tagging“ bzw. „semantic tagging“

eingegangen, wobei z.B. auch Entwicklungen wie *Text Encoding Initiative* oder *Dublin Core* zur Sprache kommen.

- Als weitere „semantic capabilities“ besprechen die Autorinnen „semantic classification“ (mit diesem Quasi-Pleonasmus ist die Zuordnung von semantischen Wissensstrukturen aus Thesauren und Klassifikationssystemen zu Synonymlisten, Topic Maps und Ontologien gemeint), „Synsets“ (Gruppierung von synonymen Begriffen, z.B. WordNet), Topic Maps und regelbasierte Systeme (rules engines).

Im dritten Hauptteil (47 Seiten) werden in vier Kapiteln Fallstudien dargestellt, die die praktische Umsetzung der zuvor behandelten theoretischen Aspekte illustrieren sollen. Dabei wird jeweils ein Informationsproblem in einer Organisation beschrieben und erläutert, wie diese Organisation Ontologien und semantische Technologien zu dessen Bewältigung einsetzt bzw. welche Pläne dort für die weitere Zukunft bestehen. Diese Darstellungen sind sehr detailliert und können hier nicht näher wiedergegeben werden. Daher sei nur angeführt, um welche Organisationen es sich handelt:

- Biogen Idec, ein Unternehmen auf biotechnologischem Gebiet (drug discovery research);
- Center on Media and Child Healthcare, eine Einrichtung der Bostoner Kinderklinik (exploring effects of media);
- Partners HealthCare Systems, ein grosses Unternehmen aus dem Gesundheitsbereich (clinical decision support);
- MINDSWAP, eine Forschergruppe der Universität Maryland (support to counter-terrorism analysts).

Der vierte Hauptteil, *Advanced topics*, ist kürzer (22 Seiten) und bietet interessierten Lesern – den anderen raten die Autorinnen, den Abschnitt zu überspringen – eine detaillierte Einführung zu den Sprachen zur Darstellung von Ontologien (XML, RDF, OWL, SKOS) und den Werkzeugen für die Erstellung von Ontologien. Dabei wird erwähnt, dass 2002 erst 56, 2004 aber bereits 94 derartige Tools angeboten wurden. Als bekanntestes wird die Stanford-Entwicklung *Protegé* erwähnt. Als allgemein bestes Tool für die Erstellung einer Ontologie bezeichnen King und Reinold das jeweils einfachste zur Lösung der betreffenden Aufgabe – eine sympathische Aussage, bei der man sich an Ockhams Rasiermesser erinnert. Des weiteren werden auch einige Kriterien zur Evaluierung solcher Werkzeuge andiskutiert.

Nur neun Seiten umfasst schliesslich der letzte Abschnitt, *Transitions to the future*. Zunächst erfolgt ein Rückbezug auf die bekannte Vision eines *Semantic Web* (Berners-Lee et al., 2001) und die dort erhobene Forderung, den Daten und Links des WWW die fehlende Semantik hinzuzufügen. Damit ist es allerdings – was kein Geheimnis ist – auch heute noch nicht weit

her. Dass Ontologien gar nicht so einfach zu erstellen sind, konzедieren die Autorinnen nun spät, aber doch. Auch die Tatsache, dass sie am besten für eher eng abgesteckte Themengebiete geeignet sind, macht ihre Anwendung auf breiter Basis problematisch. Kann es Verbindungen zwischen verschiedenen Ontologien geben? Unter der Bezeichnung „ontology mediation“ gibt es tatsächlich entsprechende Entwicklungsbestrebungen, die aber wohl noch in den Kinderschuhen stecken. Auch die Welt der semantischen Tools ist noch nicht in einem ausgereiften Stadium. Kollaborative Ansätze auf breiterer Basis sind zwar populär, leiden aber unter mangelnder Autorität. Bibliothekarisches Know-how, so die Autorinnen in der Schlusspassage, könnte hier vielleicht erfolversprechend eingesetzt werden.

Im Anhang enthält der Band noch eine Sammlung von bibliographischen und Web-Ressourcen, ein Verzeichnis der zitierten Literatur sowie ein Sachregister.

Ein kurzes Fazit: Ein lesenswertes Buch, das insbesondere jenen, die sich bislang noch nicht mit Ontologien und semantischen Techniken befasst haben, einen nützlichen ersten Einblick zu geben vermag. Nichts hingegen für Spezialisten, denn der Text kann auf den verfügbaren Seiten an keiner Stelle wirklich in die Tiefe gehen. Insbesondere die beiden ersten Hauptabschnitte sind auch als Lehrtext gut geeignet, selbst wenn an verschiedenen Stellen die Präzision zu wünschen übrig lässt. Gar zu nonchalant wird hier nämlich mit Grundbegriffen der Wissensorganisation umgegangen – z.B. wird der Informationsthesaurus nicht ausreichend vom Synonymenwörterbuch (Roget's) abgegrenzt, Klassifikationssysteme werden durchweg als Taxonomien (mit reduziertem Anspruchsniveau) verkauft. Und: zum Themenbereich „automatisches Klassifizieren“ fällt den Autorinnen lediglich das reichlich angestaubte System *Scorpion* von OCLC ein. Mit diesen kleinen Einschränkungen kann das Werk aber durchaus empfohlen werden. Der Ladenpreis ist natürlich für Studierende bzw. persönliche Käufer zu hoch, doch das Buch wird wohl ohnedies vornehmlich institutionelle Abnehmer finden.

Otto Oberhauser, Wien

Berners-Lee, T.; Hendler, J.; Lassila, O. (2001). The Semantic Web. *Scientific American magazine*. (May). Online: <http://www.scientificamerican.com/article.cfm?id=the-semantic-web> [25.05.2009]

Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge acquisition*. 5(2). 199-220.

World Wide Web Consortium. (2004). *OWL Web Ontology Language: Use cases and requirements*. Online: <http://www.w3.org/TR/2004/REC-webont-rec-20040210/> [25.05.2009]